Adaptive Sampling and Control for POMDPs: Application to Precision Agriculture

D.J. Antunes, R.M. Beumer, M.J.G. van de Molengraft, W.P.M.H. Heemels

***Abstract*— Given a partially observable Markov decision process (POMDP) with finite state, input and measurement spaces and costly measurements and control, we consider the problem of when to sample and actuate. Both sampling and actuation are modeled as control actions in a framework that encompasses both estimation and intervention problems. The process evolves freely, only driven by disturbances, in between two consecutive control actions. Control actions are assumed to reset the conditional distribution of the state given the measurements to one of a finite number of distributions. We tackle the problem of finding the times at which these control actions should take place in order to minimize an average cost that penalizes states and the number of control actions. The problem is first shown to boil down to a stopping time problem. While the latter can be solved optimally, the complexity of the optimal policy is intractable. Thus, we propose two approximate methods. The first is inspired by relaxed dynamic programming and it is within an additive cost factor of the optimal policy. The second is inspired by consistent event-triggered control and ensures that the cost is smaller than that of periodic control for the same control rate. We conclude that the latter policy can deal with large dimension problems, as demonstrated in the context of precision farming. In such a context, the proposed policies rely on a limited number of remote sensors indicating the presence of weed to decide the timings of weed removal.**

I. Introduction

There are many control and estimation applications where taking actions or intervening in an otherwise freely evolving process is costly but necessary. Thus, the times of these interventions must be carefully selected, possibly based on available process information. For instance, in remote monitoring limited sensors can work as a proxy to detect events that need to be confirmed by closer inspection and eventually handled. This is the case in precision farming, where diseases, water stress, nitrogen levels, or weed can be inferred, to some extent, from a few sensors placed in the field. However, costly farmer or (aerial) robot inspections are required for full situational awareness and handling. See [1], [2] where when to irrigate depends on soil moisture and temperature sensors, and [3] where when to apply nitrogen fertilizers depends on in-situ active-light reflectance measurements. Similar challenges arise in queuing control [4], [5], predictive maintenance [6], [7], stock trading [8], home surveillance [9], and disease outbreaks [10].

These and related problems have been studied in several fields mostly considering finite (or countable) state, input and measurement spaces, leading to partially observable Markov decision processes (POMPDs) [4]–[8]. It is well-known that these problems are intractable and thus finding appropriate (close to optimal) strategies remains an important challenge. In turn, over the past two decades, much research has been carried out in the area of event-triggered control [11] that tackles the related problem of choosing the times to sample or actuate in a control loop based on states or events rather than the usually considered periodic time-triggered sampling and control strategies. In event-triggered control, state, input and measurement spaces are typically continuous, and therefore the literature that does consider finite spaces is scarce. A line of related research proposed in [12], [13] and [14], aims at finding control policies for POMDPs considering finite spaces that depend on events, defined in terms of given state transitions; events are pre-defined, whereas in many event-triggered control papers, as in the present paper, events are also to be scheduled as a function of the process information based on an optimal criterion [15]–[18]. This latter approach is followed in a different research line proposed in [19], [20] and [21], also considering POMDPs with finite spaces. However, [19], [20] proposes to decide the next sampling or control time based on the information up to the current sampling or control time and not in between the two. This parallels self-triggered control [11], which differs from event-triggered control. In fact, event-triggered control continuously monitors the state or output of the process to decide the next sampling or control time [11], which is the approach taken here.

To be more precise, the present paper considers POMDPs with finite state, input, and measurement spaces. Control actions are used to model both costly sampling through information gathering and costly actuation through process intervention. Since these are costly, control actions only take place at some discrete times and the process evolves freely in between two consecutive control actions, possibly with inputs computed at control action times. Control actions are assumed to reset the conditional distribution of the state given the measurements to one of a finite number of distributions. In this sense, the effect of control actions is known and only *when* control actions should be enforced is to be determined. In fact, we tackle the problem of finding the times at which these control actions should take place in order to minimize an average cost that additively penalizes state configurations and the number of control actions. We show that the average cost problem can be tackled as a stopping time problem. While for this latter problem, an optimal policy can be obtained, its complexity is intractable. This motivates us to

The authors are with the Control Systems Technology Group, Department of Mechanical Engineering, Eindhoven University of Technology, the Netherlands. E-mails:{d.antunes, r.m.beumer, w.p.m.h.heemels, m.j.g.v.d.molengraft}@tue.nl. This research is part of the research program SYNERGIA (project number 17626), which is partly financed by the Dutch Research Council (NWO).

propose the two classes of approximate policies.

First, we propose a class of policies inspired by relaxed dynamic programming, see [22]. These policies guarantee a cost within an additive constant factor of the cost of the optimal policy, rather than a multiplicative factor as in original relaxed dynamic programming [22]. As explained in the sequel, this is needed for the stopping time problem at hand since the cost can be negative. While the complexity of the approximate policy is by far smaller than that of the optimal policy, and it provides nearly optimal results when the state dimension is small, it becomes impractical when the state dimension is large (see example in Section VI).

Second, inspired by [17], [18], we propose a class of so-called consistent policies that lead to a strictly smaller cost than that of periodic inspection for the same average inspection rate. The policies proposed here are different from the ones in [17], [18], both in form and in terms of their derivation, to account for the case that the state probability distribution resets to one of a set of possible distributions rather than a single one, which is a crucial assumption to capture important applications such as remote estimation.

The applicability of the results is highlighted by a numerical case study in the context of precision farming. The proposed policies rely on limited remote sensors indicating the presence of weed to decide the timings of weed removal. Due to space discretization the state dimension is rather large. Differently from relaxed dynamic programming, the policy inspired by consistent event-triggered control can handle problems with a large state dimension. This shows that ideas inspired by the event-triggered control literature can be useful to determine the times to sample and control a POMDP.

The remainder of the paper is organized as follows. Section II provides the problem formulation and some applications that can be tackled within this framework. Section III provides the optimal policy and explains the intractability issue. Sections IV and V provide the approximate policies and main results, based on relaxed dynamic programming and the consistent policies, respectively. Simulation results are discussed in Section VI and concluding remarks in Section VII. Due to space limitations the proofs of the results are omitted but can be found in the appendix.

II. Problem formulation

Consider a dynamical system with discrete state, input, and measurement spaces

$$\begin{aligned} x_{t+1} &= \underline{f}(x_t, u_t, w_t) \\ y_t &= \underline{h}(x_t, u_t, v_t) \end{aligned} \tag{1}$$

where $x_t \in \{1, 2, \dots, n\}$, $u_t \in \{1, 2, \dots, n_u\}$, $y_t \in \{1, 2, \dots, n_y\}$, $w_t \in \{1, 2, \dots, n_w\}$, $v_t \in \{1, 2, \dots, n_v\}$ are the state, control input, measurement, process disturbance input, and measurement noise input at time $t \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, respectively. The disturbance sequences $\{w_t | t \in \mathbb{N}_0\}$ and $\{v_t | t \in \mathbb{N}_0\}$ are assumed to be independent and identically distributed disturbance sequences (i.i.d.), which are also mutually independent. The fact that the output equation in (1) also depends on the control input u_t allows to tackle sensor management problems [23]. Consider also an average cost given by

$$J_s = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\underline{g}(x_t, u_t)] \tag{2}$$

Proper (ergodicity) assumptions will be given in the sequel for this limit to exist (see Assumption 2). If u_t has no restrictions then this is a standard infinite-horizon POMDP. However, in our formulation selecting or modifying u_t is costly and we impose appropriate restrictions.

Section II-A provides such a formulation and states the problem. Section II-B provides some examples of applications that are encompassed by the proposed framework.

A. Formulation with costly control rate and problem statement

The control rate is assumed to be costly and thus u_t can only be decided upon at so-called control times $s_\ell \in \mathbb{N}_0$,

$$s_{\ell+1} = s_\ell + \tau_\ell, \tag{3}$$

with $\tau_\ell \in \mathbb{N}$. For convenience, the intervals between control times are assumed to be bounded

$$\tau_\ell \le \bar{h}, \quad \forall \ell \in \mathbb{N}_0. \tag{4}$$

Since $\bar{h} \in \mathbb{N}$ can be arbitrarily large, this assumption is not very restrictive in practice. We introduce a binary control variable to indicate the time decisions of control actions

$$\sigma_t = 1, \text{ if } s_\ell = t \text{ for some } \ell$$

and $\sigma_t = 0$ otherwise. We assume that $s_0 = 0$ and, thus, $\sigma_0 = 1$. Note that $\{t \in \mathbb{N}_0 | \sigma_t = 1\} = \{s_\ell | \ell \in \mathbb{N}_0\}$. At times $t = s_\ell$, a sequence of current and future inputs $U_{0|t}$, $U_{1|t}$, …, $U_{\bar{h}-1|t}$ is computed to be applied to the system in between control times

$$u_{t+j} = U_{j|t} \text{ for } j \in \{0, \dots, \tau_\ell - 1\} \text{ when } t = s_\ell \text{ for some } \ell.$$

Often there is one decision in the set $u_t = \underline{u} \in \{1, \dots, n_u\}$ that corresponds to a free (not controlled) mode of the system and $U_{j|s_\ell} = \underline{u}$ for $j \in \{1, \dots, \tau_\ell - 1\}$. However, more general cases can be considered provided that the $U_{j|s_\ell}$ satisfy Assumption 3 below.

Let $\tilde{p}_0$ denote the initial state probability distribution

$$\tilde{p}_0 = \begin{bmatrix} \tilde{p}_{0,1} & \tilde{p}_{0,2} & \dots & \tilde{p}_{0,n} \end{bmatrix}^\intercal$$

with $\tilde{p}_{0,i} = \text{Prob}[x_0 = i]$ and $\tilde{p}_0 \in \mathcal{P}_n := \{p = [p_1 \dots p_n]^\intercal | 1_n^\intercal p = 1, p_i \ge 0, \forall i\}$, where 1_n denotes a column vector with n entries equal to one. Let also $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{y_t, \sigma_{t-1}, u_{t-1}\}$ for $t \in \mathbb{N}$ with $\mathcal{I}_0 = \{\tilde{p}_0\} \cup \{y_0\}$ denote the information available for decisions up to time t and

$$p_{t|t} = \begin{bmatrix} p_{t|t,1} & p_{t|t,2} & \dots & p_{t|t,n} \end{bmatrix}^\intercal$$

denote the probability distribution of the state x_t given the information set $\mathcal{I}_t$, i.e.,

$$p_{t|t,i} = \text{Prob}[x_t = i | \mathcal{I}_t].$$

with $p_{t|t} \in \mathcal{P}_n$. Let $p_{t+1|t}$ be defined similarly but with

$$p_{t+1|t,i} = \text{Prob}[x_{t+1} = i|\mathcal{I}_t].$$

A crucial assumption is that either $p_{t|t}$ or $p_{t+1|t}$ belongs to a known set of b possible distributions, denoted by $\rho_1, \dots, \rho_b$, when actuation is computed (at control times). Formally:

Assumption 1: One of the following conditions holds, for every $\ell \in \mathbb{N}_0$,

$$\text{(i) } p_{s_\ell|s_\ell} \in \{\rho_1, \dots, \rho_b\}, \tag{5a}$$

$$\text{(ii) } p_{s_\ell+1|s_\ell} \in \{\rho_1, \dots, \rho_b\}. \tag{5b}$$

□

Assumption 1(i) captures applications where the state becomes either known or has a known probability distribution at time t (through map $\underline{f}$) when a control intervention on the process is carried out at time t, while Assumption 1(ii) captures applications where the control actions at time t influence measurements at time t through map $\underline{h}$. Two examples are given in Section II-B.

Let $\phi_\ell \in \{1, \dots, b\}$ be such that $p_{s_\ell|s_\ell} = \rho_{\phi_\ell}$ or $p_{s_\ell+1|s_\ell} = \rho_{\phi_\ell}$, when Assumptions 1(i), 1(ii) hold respectively (if both hold either choice for ϕ_ℓ can be picked). We can define a Markov chain with b states and transition probability matrix R with entries (i) $R_{ij} = \text{Prob}[\phi_{t+1} = i|\phi_t = j]$. This Markov chain is assumed to be ergodic, meaning that it is aperiodic and irreducible. Thus, it has a stationary probability distribution.

Assumption 2: There exists a unique $a \in \mathcal{P}_b$ such that $a = Ra$.

□

Due to this assumption and since we are interested in an average cost (2) we can without loss of generality assume that the initial distribution of ϕ_0 is a, that is $\text{Prob}[\phi_0 = i] = a_i$, $i \in \{1, \dots, b\}$.

A third assumption imposes that the control sequence $U_{j|s_\ell}$ only depends on ϕ_ℓ.

Assumption 3: $U_{j|s_\ell} = \theta(j, \phi_\ell)$ for every $j \in \{0, \dots, \bar{h} - 1\}$, every $\ell \in \mathbb{N}_0$, and for some function θ.

□

These assumptions are motivated by and met in the applications discussed in the sequel.

Note that due to these assumptions in between control times the system evolves freely according to

$$x_{t+1} = f(x_t, \zeta_t, \phi_\ell, w_t), \quad s_\ell \le t \le s_{\ell+1} - 1,$$

where

$$\zeta_t = t - s_{\bar{\ell}(t)}, \quad \bar{\ell}(t) = \max\{\ell | s_\ell \le t\},$$

and $f(x_t, \zeta_t, \phi_\ell, w_t) = \underline{f}(x_t, \theta(\zeta_t, \phi_\ell), w_t)$. Likewise, the average cost can be written as

$$J_s = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{\ell=0}^{L(T)-1} \sum_{t=s_\ell}^{s_{\ell+1}-1} g(x_t, \zeta_t, \phi_\ell) + \sum_{t=s_{L(T)}}^{T-1} g(x_t, \zeta_t, \phi_{L(T)}) \right], \tag{6}$$

where $L(T) = \max\{\ell | s_\ell \le T - 1\}$, $g(x_t, \zeta_t, \phi_\ell) = \underline{g}(x_t, \theta(\zeta_t, \phi_\ell))$, and the output as

$$y_t = h(x_t, \zeta_t, \phi_\ell, v_t), \quad s_\ell \le t \le s_{\ell+1} - 1,$$

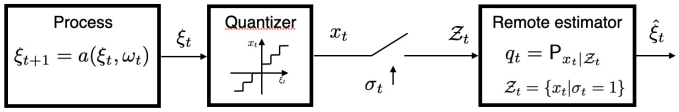

Fig. 1: Remote estimation: The quantized state of a process is sent through a costly link to a remote estimator that computes the quantized state distribution given the previous received states, denote by $\mathsf{P}_{x_t|\mathcal{Z}_t}$, and computes an estimate $\hat{\xi}_t$ of ξ_t according to (10).

where $h(x_t, \zeta_t, \phi_\ell, v_t) = \underline{h}(x_t, \theta(\zeta_t, \phi_\ell), v_t)$.

The fact that the actuation is costly is handled by defining a cost that penalizes the average rate of control actions $J_c := \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\sigma_t]$. Then, one can see the minimization of

$$J_{av} = J_s + \delta J_c \tag{7}$$

over different values of δ as finding Pareto optimal policies. Another interpretation for the cost J_{av} is that there is an actual actuation cost and noticing that the running cost for J_{av} is $\underline{g}(x_t, u_t) + \delta\sigma_t$. The goal is to find a policy

$$\sigma_t = \mu_t(\mathcal{I}_t) \tag{8}$$

for $t \in \mathbb{N}_0$, that minimizes J_{av}.

B. Examples of applications

1) Remote estimation with costly transmission: Consider a process described by a special case of (1) that does not depend on a control input and for which the full state is available

$$x_{t+1} = f(x_t, w_t), \quad y_t = x_t. \tag{9}$$

The dynamical model can equivalently be described by a transition matrix P with entries $P_{ij} = \text{Prob}[x_{t+1} = i|x_t = j]$, assumed to be ergodic. Model (9) results from quantizing a process $\xi_{t+1} = a(\xi_t, \omega_t)$ with $\xi_t \in \mathbb{R}^n$ and ω_t i.i.d. disturbances, and state x_t labels one of n representative values $\underline{\xi}^i$ of the quantized state variable ξ_t. This is illustrated in Figure 1. Thus, there exists a labelling map π such that $\underline{\xi}^i = \pi(i)$. The state is known to an agent who wishes to send it to a remote estimator. Control times s_ℓ are understood here in a broad sense as the times at which information is sent to the remote estimator, and, as before, are indicated by $\sigma_t \in \{0, 1\}$ determining when information is sent ($\sigma_t = 1$) or not ($\sigma_t = 0$). Transmissions are assumed to be expensive, e.g., due to battery limitations on the sensor side. On the remote side, an estimator obtains $q_t = \begin{bmatrix} q_{t,1} & \dots & q_{t,n} \end{bmatrix}^\intercal$ and $q_{t,i} = \text{Prob}[x_t = i|\mathcal{Z}_t]$ with $\mathcal{Z}_t := \{y_\ell | 0 \le \ell \le t, \sigma_\ell = 1\}$, and $q_t \in \mathcal{P}_n$ and computes

$$\hat{\xi}_t := \mathbb{E}[\pi(x_t)|\mathcal{Z}_t] = \sum_{i=1}^{n} \pi(i) q_{t,i} \tag{10}$$

Note that, assuming $\sigma_0 = 1$,

$$q_t = \begin{cases} \delta_{x_t}, \text{ if } \sigma_t = 1, \\ P^{\zeta_t} \delta_{x_{t-\zeta_t}}, \text{ if } \sigma_t = 0, \end{cases} \tag{11}$$

where δ_i is the column vector of zeros except at position i where it equals 1. The running cost of an average cost penalizes through the euclidean norm the difference between the state and the estimated state

$$\|\pi(x_t) - \hat{\xi}_t\|^2. \tag{12}$$

While (9) does not depend on the control input u_t we can use the control input as a modeling variable to ensure we can write (2) with running cost (12). In fact, we can define

$$u_{1,t} = U_{\zeta_t|s_\ell} = x_{s_\ell}, \quad u_{2,t} = \zeta_t, \text{ for } s_\ell \le t \le s_{\ell+1} - 1,$$

and $u_t \in \{1, \dots, n_u\}$, with $n_u = n\bar{h}$ to assign a unique label to the pair $(u_{1,t}, u_{2,t}) \in \{1, \dots, n\} \times \{1, \dots, \bar{h}\}$ Then (12) can be written as $\underline{g}(x_t, u_t)$ and (10), (11) are also functions of the state and control input since, when $\sigma_t = 0$, $q_t = P^{u_{2,t}} \delta_{u_{1,t}}$ and $\sigma_t = 1$, $q_t = \delta_{x_t}$. Moreover, $p_{s_\ell|s_\ell} = \delta_{x_{s_\ell}} \in \{\delta_1, \dots, \delta_n\}$ belongs to a finite set, $U_{j|s_\ell}$ are functions of $\phi_\ell = x_{s_\ell}$ and j and ergodicity of R follows from ergodicity of P. The goal is to find a policy (8) for when to apply control actions (send remote data) in order to minimize (7) where the running cost in J_s is given by (12).

2) Costly interventions based on limited data: Consider a set of discrete N interdependent states $x_i \in \{1, \dots, n_i\}$ evolving in a free, or non-controlled, fashion according to

$$x_{i,t+1} = f_i(x_{1,t}, \dots, x_{N,t}, w_{i,t}), \quad i \in \{1, \dots, N\}, \tag{13}$$

when there are no control interventions ($\sigma_t = 0$), where the disturbance inputs $w_{i,t}$ live in a finite set. At control or intervention times ($\sigma_t = 1$) the state is reset to

$$x_{1,t+1} = \alpha_1, \ x_{2,t+1} = \alpha_2, \ \dots \ x_{N,t+1} = \alpha_N \tag{14}$$

and a fixed cost δ is paid for each intervention. Only a subset of states is measured by M sensors. Each sensor $j \in \{1, \dots, M\}$, depends on a subset of n_j states $\mathcal{L}_j = \{l_1, \dots, l_{n_j}\}$ with $l_j \in \{1, \dots, N\}$,

$$y_j = h_j(x_{l_1}, \dots, x_{l_{n_j}}, v_j) \tag{15}$$

where $j \in \{1, \dots, M\}$, and where the measurements $y_j \in \{1, \dots, n_{y,j}\}$ might be corrupted by noise v_j, also living in a finite set. The running cost of an average cost is given by

$$g_A(x_{1,t}, \dots, x_{N,t}) = \sum_{i=1}^{N} g_i(x_{i,t}) \tag{16}$$

In the context of precision agriculture (see Section VI for more details), each $x_{i,t}$ might be a binary variable representing if there is weed ($x_{i,t} = 2$) in a given subarea of a field at time t or not ($x_{i,t} = 1$); $x_{i,t}$ depends on neighboring states and only a subset of subareas can be measured. At intervention times s_ℓ the weed is completely removed, setting all the states to $x_{i,s_\ell+1} = 1$, $i \in \{1, \dots, N\}$. For some constant d, representing the cost of having weed between t and $t+1$,

$$g_i(2) = d \text{ and } g_i(1) = 0 \quad \forall i \in \{1, \dots, N\}. \tag{17}$$

We can find single variables $x_t \in \{1, \dots, n\}$, $y_t \in \{1, \dots, n_y\}$, $n = n_1 \times \cdots \times n_N$, $n_y = n_{y,1} \times \cdots \times n_{y,M}$ to label all possible state and output combinations and write the problem in the canonical formula described above, see Section VI. Here $p_{t+1|t} = \delta_\alpha$, corresponding to $x_{t+1} = \alpha$, where $\alpha \in \{1, \dots, n\}$ is the label corresponding to (14), and Assumption 1 is trivially met since R corresponds to a Markov chain with just one state. The control input u_t can be used to model the process evolution and in particular the state reset (14), but it can also be omitted. The goal is to find a policy for σ_t as a function of the information set $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{y_t, \sigma_{t-1}\}$ for $t \in \mathbb{N}$, with $\mathcal{I}_0 = \{y_0\}$ to minimize the average cost (7) with the proposed running cost.

III. Optimal policy

We start by providing a result that allows for simplifying the problem from an average cost problem to a stopping time problem. The stopping time problem is the following: consider system (1), which now needs to be considered only in the interval $t \in \{0, 1, \dots, \bar{h}\}$. The information available to make a decision at time $t \in \{1, \dots, \bar{h}\}$ on either $s_1 = \tau_0 \in \{1, \dots, \bar{h}\}$ is equal to t or larger, is summarized in $p_{0|0} = \rho_{\phi_0}$, if (5a) holds, or in $p_{1|0} = \rho_{\phi_0}$, if (5b) holds, and in the measurements y_1, y_2, …, y_t. Thus, we define the information set

$$\mathcal{H}_t^0 = \{\phi_0\} \cup \{y_\ell | \ell \in \{1, \dots, t\}\}.$$

Consider now the following optimal stopping time problem first for $\ell = 0$ and $s_0 = 0$:

$$J_{\text{stop}} = \min_{\tau_\ell} \frac{1}{\mathbb{E}[\tau_\ell]} \left(\mathbb{E}[\sum_{t=0}^{\tau_\ell - 1} g(x_{s_\ell+t}, \zeta_{s_\ell+t}, \phi_{s_\ell})] + \delta \right) \tag{18}$$

where τ_0 is a stopping time with respect to the filtration corresponding to the information set $\mathcal{H}_t^0$. In other words, the event $[\tau_0 = m]$ is a function of $\mathcal{H}_m^0$. Having defined τ_0 we can similarly define stopping times τ_ℓ with respect to the filtration corresponding to the information set $\mathcal{H}_t^\ell = \{\phi_\ell\} \cup \{y_{s_\ell+1}, y_{s_\ell+2}, \dots, y_{s_\ell+t}\}$ and consider an identical stopping time problem to (18) for a general ℓ. These problems are identical due to Assumption 1, 2, 3 as stated next. These stopping times τ_ℓ define the control times according to (3).

Lemma 1: Suppose that Assumptions 1, 2, 3 hold. Then the optimal stopping time policies for problems (18) are identical in the sense that they take the form

$$\tau_\ell = \min\{r \ge 1 | \sigma_{s_\ell+r} = 1\}$$

$$\sigma_t = \xi_{\zeta_t}(H_{\zeta_t}^\ell) \quad \ell \in \mathbb{N}_0 \tag{19}$$

for the same functions ξ_i, $i \in \{0, 1, \dots, \bar{h} - 1\}$. Moreover, (19) is also an optimal policy for the average cost problem of minimizing (7) and $J_{\text{av}} = J_{\text{stop}}$. Furthermore, the optimal policy for problem (18) can be obtained by solving the following stopping time problem

$$\min_{\tau_0} \mathbb{E}[\sum_{t=0}^{\tau_0 - 1} (g(x_t, \zeta_t, \phi_0) - \beta)] + \delta \tag{20}$$

where $\beta \in \mathbb{R}_{\geq 0}$ is the largest value for which the optimal solution to this new stopping problem (20) results in a zero cost and is given by $\beta = J_{\text{av}}$.

□

In the proof of Lemma 1 we show that (20) is a non-increasing continuous function of β and thus has a unique zero (since for $\beta = 0$ it is positive and for β very large it is negative).

The optimal policy and cost for problem (20) for a given β can be obtained by the dynamic programming algorithm as shown next. The search for β is discussed at the end of the section. Let $q_t = \begin{bmatrix} q_{t,1} & \dots & q_{t,n} \end{bmatrix}^\intercal$ with $q_{t,i} = \text{Prob}[x_t = i|\mathcal{H}_t^0]$ and

$$\bar{g}_t = \begin{bmatrix} g(1,t,\phi_0) & \dots & g(n,t,\phi_0) \end{bmatrix}^\intercal, t \in \{0,1,\dots,\bar{h}-1\},$$

where the dependence of $\bar{g}_t$ on the given ϕ_0 is omitted, as well as in other functions in the sequel, both for notation convenience and since often g does not depend on ϕ_0. Then, one should iterate for $t \in \{\bar{h}-1,\dots,0\}$

$$\begin{aligned} J_{\bar{h}}(q_{\bar{h}}) &= \delta, \\ J_t(q_t) &= \min\{\delta, q_t^\intercal \bar{g}_t - \beta + \mathbb{E}[J_{t+1}(q_{t+1})|\mathcal{H}_t^0]\} \end{aligned}$$

and the optimal policy is

$$\tau_\ell = \min\{r \in \{1,\dots,\bar{h}\}|\sigma_{s_\ell + r} = 1\}$$

$$\sigma_t = \begin{cases} 1 \text{ if } J_{\zeta_t}(q_{s_\ell+\zeta_t}) = \delta \text{ or if } \zeta_t = \bar{h} \\ 0 \text{ otherwise.} \end{cases}$$

Note that q_t can be iterated with the Bayes' filter

$$q_{t+1} = \frac{1}{\alpha(y_{t+1})} D(y_{t+1}) P q_t$$

with $P_{ij} = \text{Prob}[x_{t+1} = i|x_t = j]$ and

$$D(y_{t+1}) = \text{diag}([R_{y_{t+1},1} \dots R_{y_{t+1},n}])$$

with $R_{\ell,j} = \text{Prob}[y_t = \ell|x_t = j]$, for $\ell \in \{1,\dots,m\}$, $j \in \{1,\dots,n\}$ and

$$\alpha(\ell) = \text{Prob}[y_{t+1} = \ell|\mathcal{I}_t] = \sum_{j=1}^{n} R_{\ell,j} \bar{r}_{t,j}.$$

with $\bar{r}_t = [\bar{r}_{t,1} \ \dots \ \bar{r}_{t,n}]^\intercal$, $\bar{r}_t = P q_t$. The filter is initialized with $q_0 = \rho_{\phi_0}$ if (5a) holds and $\bar{r}_0 = \rho_{\phi_0}$ if (5b) holds.

It is well known that (see, e.g., [22])

$$J_t(q_t) = \min_{c \in \mathcal{J}_t} c^\intercal q_t$$

However, the size of set $\mathcal{J}_t$, denoted by $|\mathcal{J}_t|$, grows as

$$|\mathcal{J}_t| = 1 + |\mathcal{J}_{t+1}|^{n_y} \tag{21}$$

leading to a computationally intractable method to find the optimal policy. We will briefly show this as the derivation is important in the sequel. Assume that $J_{t+1}(q_{t+1}) = \min_{c \in \mathcal{J}_{t+1}} c^\intercal q_{t+1}$, which is true for $t = \bar{h}-1$ as we can write $J_{\bar{h}}(q_{\bar{h}}) = \beta = c^\intercal q_{\bar{h}}$ with $c = \beta 1_n$ since $1_n^\intercal p_{\bar{h}} = 1$. To obtain an expression for $J_t(q_t)$ we need to compute

$$\begin{aligned} \mathbb{E}[J_{t+1}(q_{t+1})|\mathcal{H}_t^0] &= \sum_{\ell=1}^{n_y} \mathbb{E}[J_{t+1}(q_{t+1})|y_{t+1} = \ell, \mathcal{I}_t]\alpha(\ell) \\ &= \sum_{\ell=1}^{n_y} J_{t+1}(\frac{D(\ell)}{\alpha(\ell)} P q_t)\alpha(\ell) \\ &= \sum_{\ell=1}^{n_y} \min_{c \in \mathcal{J}_{t+1}} c^\intercal (D(\ell) P q_t) \end{aligned}$$

Note that $\delta = d_2^\intercal q_t$, $\quad q_t^\intercal \bar{g}_t - \bar{\beta} = d_{1,t}^\intercal q_t$ for $d_2 = \delta 1$ and $d_{1,t} = \bar{g}_t - \bar{\beta} 1$ so that

$$\begin{aligned} J_t(q_t) &= \min(d_2^\intercal q_t, d_{1,t}^\intercal q_t + \sum_{\ell=1}^{n_y} \min_{c \in \mathcal{J}_{t+1}} c_i^\intercal (D(\ell) P q_t)) \\ &= \min_{c \in \mathcal{J}_t} c^\intercal q_t \end{aligned}$$

where $\mathcal{J}_t = \mathcal{J}_t^1 \cup \{d_2\}$ and

$$\begin{aligned} \mathcal{J}_t^1 = \{d_{1,t} + \sum_{\ell=1}^{n_y} P^\intercal D(\ell)^\intercal \bar{c}_{j_\ell} | \bar{c}_{j_\ell} \in \mathcal{J}_{t+1}, \\ j_1,\dots,j_{n_y} \in \{1,\dots,|\mathcal{J}_{t+1}|\}\} \end{aligned}$$

Note that indeed (21) holds.

When $\bar{g}_t$ depends on $\phi_0 = i$ so will the cost to go and now we stress this by denoting the cost to go at time $t = 0$ by $J_0^{i,\beta}(q_t)$ where the dependence on β is also added. The cost (20) is then equal to $\sum_{i=1}^{b} a_i J_0^{i,\beta}(\rho_i)$. Thus, one needs to find β for which this cost is zero. To this end we can simply run a bisection algorithm as this is a monotone and continuous function of β.

IV. Relaxed dynamic programming policy

The idea of relaxed dynamic programming [22] is to find simpler functions to approximate $J_t(q_t)$. Here we consider the following functions $V_{\bar{h}}(q_{\bar{h}}) = \beta$ and

$$V_t(q_t) = \min_{c \in \mathcal{V}_t} c^\intercal q_t, \quad t \in \{0,1,\dots,\bar{h}-1\} \tag{22}$$

where $\mathcal{V}_t \subseteq \mathcal{J}_t$ can be seen as a pruned version of $\mathcal{J}_t$. We will provide a procedure to obtain this pruned set in such a way that

$$J_t(q_t) \leq V_t(q_t) \leq J_t(q_t) + \epsilon(\bar{h}-t), \ t \in \{0,1,\dots,\bar{h}-1\} \tag{23}$$

so that $V_t(q_t)$ is always within an additive factor $\epsilon(\bar{h}-t)$ of the optimal policy and the resulting policy as well. Although the original idea of relaxed dynamic programming considered a multiplicative factor for the guarantees, i.e., $J_t(q_t) \leq V_t(q_t) \leq (1+\epsilon)J_t(q_t)$, here an additive factor is chosen for two reasons. First, in the present application J_t and V_t take in general negative values (due to subtracting β from the running cost (20)), thus the multiplicative bound makes no sense. Second in the present application $J_t(q_t) \leq \delta$ for every t and q_t so that it is easy to avoid scaling issues.

Towards this, let us define

$$J_t^\epsilon(q_t) = \min\{\delta + \epsilon, (q_t^\intercal \bar{g} - \beta + \epsilon) + \mathbb{E}[V_{t+1}(q_{t+1})|\mathcal{H}_t]\}$$

Using similar steps to the ones used in the previous section we can conclude that

$$J_t^\epsilon(q_t) = \min_{c \in \mathcal{J}_t^\epsilon} c^\intercal q_t$$

where

$$\mathcal{J}_t^\epsilon = \{d_1 + \epsilon 1 + \sum_{\ell=1}^{n_y} P^\intercal D(\ell)^\intercal \bar{c}_{j_\ell} | \bar{c} \in \mathcal{V}_{t+1}, \\ j_1, \ldots, j_p \in \{1, \ldots, |\mathcal{V}_{t+1}|\}\} \cup \{d_2 + \epsilon 1\}$$

Function J_t^ϵ coincides with J_t when $\epsilon = 0$. However, this function is defined with an extra cost term for the running cost when $\epsilon > 0$. Given a $\tilde{c}^\epsilon \in \mathcal{J}_t^\epsilon$, consider a corresponding $\tilde{c}^0$ from the set $\mathcal{J}_t^\epsilon|_{\epsilon=0}$ defined as above.

At each time t, the set $\mathcal{V}_t$ is a pruned version of the set $\mathcal{J}_t$. To facilitate the pruning operation, which is carried out iteratively, it is wise to define a heuristic function $H(c)$ which assigns a score to the elements c of a given set (say J_t^ϵ). The higher the value of $H(c)$ the larger the belief that c can be pruned.

Relaxed Dynamic Programming procedure:

1) Initialize $\mathcal{V}_t$ as empty or with some members c for which $c^\intercal \xi_i$ is minimal for some ξ_i.
2) Take the element $\tilde{c}^\epsilon$ in $\mathcal{J}_t^\epsilon \setminus \mathcal{V}_t$ with the smallest H and check if it satisfies

$$\min_{c \in \mathcal{V}_t} c^\intercal q \leq \tilde{c}^{\epsilon\intercal} q \quad \forall q \in \mathcal{P}_n. \tag{24}$$

3) If (24) is not satisfied, then add $\tilde{c}^0$ to $\mathcal{V}_t$. If there are no more elements in $\mathcal{J}_t^\epsilon$, then stop, otherwise go to step 2.

The heuristic used evaluates $c^\intercal \xi_i$ for each of n_c members c, and for s fixed values ξ_i, $i \in \{1, \ldots, s\}$ and assigns a reward $r_i \in \{1, \ldots, n_c\}$ to each member corresponding to the ranking in the ith ordered set according to $c^\intercal \xi_i$. If the ranking of a given c is one for some i, it is immediately added according to step 1). Otherwise $H(c) = \sum_{i=1}^s r_i$.

The next result states that indeed (23) is met with this procedure.

Lemma 2: Let V_t be defined by (22) with the set $\mathcal{V}_t$ obtained from the relaxed dynamic programming procedure. Then (23) holds.

□

The following test provides a sufficient condition to test if (24) holds and if (24) is replaced by this test the same guarantees can also be given: if there exists $\alpha_i \geq 0$ with $\sum_{i=1}^{|\mathcal{V}_t|} \alpha_i = 1$ such that, with $c_i \in \mathcal{V}_t$, $\sum_{i=1}^{|\mathcal{V}_t|} \alpha_i c_i \leq \tilde{c}$ then (24) holds. To test this latter condition we can simply test the feasibility of the simplex

$$A\alpha \leq b \tag{25}$$

with

$$A = \begin{bmatrix} C \\ -I \\ 1^\intercal \\ -1^\intercal \end{bmatrix}, \quad b = \begin{bmatrix} \tilde{c} \\ 0 \\ 1 \\ -1 \end{bmatrix}, \quad C = \begin{bmatrix} c_1 & \ldots & c_{|\mathcal{V}_t|} \end{bmatrix}.$$

V. Consistent policy

We start by providing a result stating the performance of periodic control.

Lemma 3: Suppose that Assumptions 1, 2, 3 hold and that $\tau_\ell = h$, $\forall \ell \in \mathbb{N}_0$, are constant, corresponding to periodic control. Then $J_{\text{av}} := J_s + \delta J_p$ with

$$J_s = \eta_h := \sum_{i=1}^b \nu_{h,i} a_i, \quad J_p = \frac{1}{h} \tag{26}$$

where

$$\nu_{h,i} := \begin{cases} \frac{1}{h}(\sum_{\ell=0}^{h-1} \bar{g}_\ell^\intercal P^\ell)\rho_i, \text{ if (5a) holds} \\ \frac{1}{h}(\sum_{\ell=0}^{h-2} \bar{g}_{\ell+1}^\intercal P^\ell + g_0^\intercal P^{\ell-1})\rho_i, \text{ if (5b) holds} \end{cases}$$

□

The proposed policy yields a better trade-off between average number of actions and average cost in the sense that if we define the curve

$$(s, J_{\text{per}}(s))$$

with

$$J_{\text{per}}(s) = \eta_r + (\eta_{r+1} - \eta_r)(s - r) \text{ if } s \in [r, r+1) \tag{27}$$

then

$$(h_{\text{av,per}}, J_{\text{av}})$$

with $h_{\text{av,per}} = \mathbb{E}[\tau_0] = 1/J_c$ the average inter-control time and J_{av} the cost of that policy, is below this curve. We call this a *consistent* policy. We propose such a consistent policy inspired by [17], [18] but different both in terms of form and in terms of derivation.

The proposed policy is defined as follows:

$$\tau_\ell = \min\{r \in \{1, \ldots, \bar{h}\} | \sum_{t=s_\ell}^{s_\ell + r} \bar{g}_t^\intercal p_{t|t} > -\delta + \omega_{c,\phi_\ell} r\} \tag{28}$$

for $\omega_{c,i} < \omega_{m,i}$ where

$$\omega_{m,i} = \min\{\nu_{r,i} + \delta \frac{1}{r} | r \in \{1, \ldots, \bar{h}\}\}.$$

Intuitively, assuming for simplicity $b = 1$ and $\ell = 0$, if we knew the state and if we could make sure that

$$\sum_{t=0}^{\tau_0 - 1} g(x_t, t, 1) + \delta - \omega_{c,1} \tau_0 \leq 0$$

we would obtain also that

$$\mathbb{E}[\sum_{t=0}^{\tau_0 - 1} g(x_t, t, 1) + \delta - \omega_{c,1} \tau_0] \leq 0,$$

which would imply that

$$\frac{1}{\mathbb{E}[\tau_0]} \mathbb{E}[\sum_{t=0}^{\tau_0 - 1} g(x_t, t, 1)] + \delta \frac{1}{\mathbb{E}[\tau_0]} \leq \omega_{c,1} < \omega_{m,1}$$

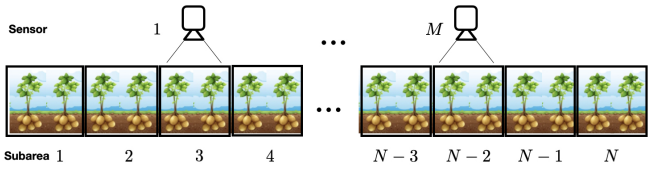

Fig. 2: Precision agriculture setting. A strip of potato crops is divided into N subareas. Weed appears and spreads along this strip. A set of M sensors measure with some fault probability whether there is weed or not in a subset of subareas. The proposed policies allow to determine when a weed removal intervention is needed.

meaning that we could ensure that such a policy would yield a better trade-off than any periodic policy. Although the state is not available, this policy still ensures this property by replacing the term $g(x_t, t, \phi_0)$ by its expected value given the information up to time t, $\bar{g}_t^{\mathsf{T}} q_t$, and by taking into account the initial condition.

A limitation is that if δ is large and $w_{m,i}$ is small the policy might lead to many control actions. Since δ can be seen as a parameter that sets the desired trade-off in the Pareto optimal curve it can be set to zero. The only restriction then comes from $\omega_{m,i}$. The consistency property holds for any choice of δ and $\omega_{m,i}$.

Theorem 1: Suppose that Assumptions 1, 2, 3 hold. Let J_π be the cost J_{av} of the proposed policy (33) and let $\bar{\tau}_\pi = \mathbb{E}[\tau_{0,\alpha}] = 1/J_c$ be the average inter-control time. Then,

$$J_\pi < \bar{J}_{\text{per}}(\bar{\tau}_\pi). \tag{29}$$

That is, the policy is consistent.

□

VI. Application in precision agriculture

Consider the precision farming setting of Section II-B.2 and assume that the field is a strip of potato crops divided in N subareas as depicted in Figure 2. The state $x_t \in \{1, \dots, 2^N\}$ labels all possible states of the N weed indicator variables $x_{t,i} \in \{1, 2\}$. At initialization and directly after the interventions at times s_ℓ, the whole field has no weed. Weed can simply appear in a given subarea or spread from one of the neighboring subareas, which is summarized by

$$\text{Prob}[x_{t+1,l} = 2 | x_{t,l-1} = a, x_{t,l} = 1, x_{t,l+1} = b] = \begin{cases} r_{11}, & \text{if } a = 1, b = 1 \\ r_{21}, & \text{if } a = 2, b = 1 \\ r_{12}, & \text{if } a = 1, b = 2 \\ r_{22}, & \text{if } a = 2, b = 2 \end{cases}$$

when $l \notin \{1, N\}$, with $0 < r_{11} < r_{21} = r_{12} < r_{22} < 1$ and $\text{Prob}[x_{t+1,1} = 2 | x_{t,2} = i, x_{t,1} = 1] = \text{Prob}[x_{t+1,N} = 2 | x_{t,N-1} = i, x_{t,N} = 1] = r_l$, $i \in \{1, 2\}$, with $0 < r_1 = r_{11} < r_2 = r_{21} < 1$. When a subarea has weed, it remains until an intervention, $\text{Prob}[x_{t+1,i} = 2 | x_{t,i} = 2] = 1$ for every i. From these assumptions (13) can be computed or equivalently the transition probability matrix $P_{ij} = \text{Prob}[x_{t+1} = i | x_t = j]$ can be computed, as follows. First, let $\pi_j(i) = x_j$ extract the value of the indicator state $j \in \{1, \dots, N\}$ for a given state $i \in \{1, \dots, 2^N\}$ and let $\mathcal{N}(i, t) = (x_{t,i-1}, x_{t,i+1})$, $\mathcal{N}(1, t) = x_{t,2}$, $\mathcal{N}(N, t) = x_{t,N-1}$ be the neighboring indicator states. Then $P_{ij} = 0$ if there exists ℓ such that $\pi_\ell(i) = 1$ and $\pi_\ell(j) = 2$ and, otherwise,

$$P_{ij} = \left(\prod_{\ell \in \mathcal{M}_{i,j,2}} r_{\mathcal{N}(\ell,t)} \right) \left(\prod_{\ell \in \mathcal{M}_{i,j,1}} (1 - r_{\mathcal{N}(\ell,t)}) \right)$$

with $\mathcal{M}_{i,j,\kappa} = \{\ell | \pi_\ell(j) = 1, \pi_\ell(i) = \kappa\}$, $\kappa \in \{1, 2\}$.

Each sensor $i \in \{1, \dots, M\}$ can measure several subareas, according to (15), but for simplicity here it is assumed it only measures one subarea $\ell_i \in \{1, \dots, N\}$. Thus $y_{t,i} \in \{1, 2\}$ and an error probability is assumed

$$\text{Prob}[y_{t,i} = 2 | x_{t,\ell_i} = 1] = \text{Prob}[y_{t,i} = 1 | x_{t,\ell_i} = 2] = e_p.$$

From these assumptions, (15) can be computed or equivalently $R_{ij} = \text{Prob}[y_t = i | x_t = j]$. Let $\chi_j(i) = y_j$ extract the value of the subarea measurement $j \in \{1, \dots, M\}$ for a given measurement $i \in \{1, \dots, 2^M\}$ and let s_{ij} be the number of subarea measurements associated with $y_t = i$ that provide a correct estimate for the corresponding subarea state associated with $x_t = j$, i.e., $s_{ij} = |\{\iota | \chi_\iota(i) = \pi_{\ell_\iota}(j)\}|$. Then

$$R_{ij} = (1 - e_p)^{s_{ij}} e_p^{M - s_{ij}}.$$

The running cost in (16), (17) can be written as the running cost in (18) and boils down to

$$g(i, t, 0) = d \begin{bmatrix} \pi_1(i) - 1 & \dots & \pi_N(i) - 1 \end{bmatrix} 1_N. \tag{30}$$

We start by considering a very special case with just one field $N = 1$, $x_t \in \{1, 2\}$, one measurement $y_t \in \{1, 2\}$ and

$$P = \begin{bmatrix} 1 - r_{11} & 0 \\ r_{11} & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 1 - e_p & e_p \\ e_p & 1 - e_p \end{bmatrix},$$

where $P_{ij} = \text{Prob}[x_{t+1} = i | x_t = j]$, $R_{ij} = \text{Prob}[y_t = i | x_t = j]$ $i, j \in \{1, 2\}$. This simple case allows one to still compute the optimal policy and understand how far, for this example, is the cost of the consistent policy. The parameters considered are $\bar{h} = 10$, $d = 1$, $r_{11} = 0.1$, $e_p = 0.2$. Figure 3 shows the results of periodic control, optimal policy (or relaxed dynamic programming with $\epsilon = 0$) considering $\delta \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5\}$ and of the consistent policy with $\delta = 0$ and

$$\omega_{m,1} \in \{0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.13, 0.15, 0.17, 0.2, 0.3, 0.4\}.$$

As it can be seen, the optimal policy yields a significant reduction of cost J_s, when the running cost is (30), for the same average intervention interval with respect to periodic control. The consistent policy uses the same parameters and provides results close to optimal.

We now consider a more realistic example with $N = 12$ subareas, $M = 3$ sensors $l_1 = 3$, $l_2 = 6$, $l_3 = 9$, still with $d = 1$, and parameters $r_{11} = 0.02$, $r_{21} = 0.2$, $r_{22} = 0.4$, $e_p = 0.02$, $\bar{h} = 20$. For this example the number of states

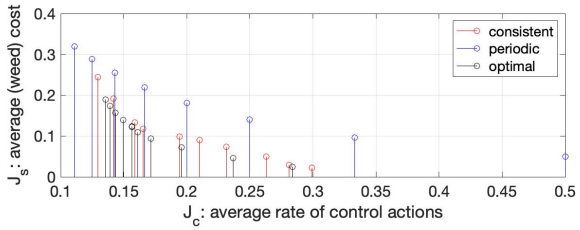

Fig. 3: Average (weed) cost J_s versus average rate of control actions J_c for three policies: periodic, optimal and consistent for a simple example with $N = M = 1$

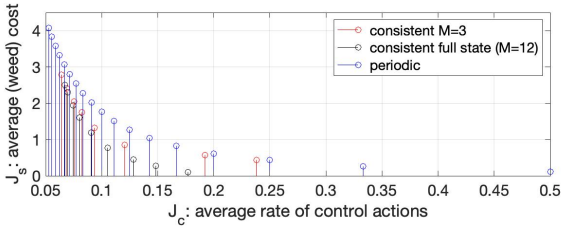

Fig. 4: Average (weed) cost J_s versus average rate of control actions J_c when $N = 12$ for periodic control, consistent with $M = 3$ sensors and full state feedback $M = 12$

is $n = 2^{12} = 4096$. We can no longer apply relaxed dynamic programming since solving (25) with $c \in \mathbb{R}^{4096}$ is computationally hard. However, we can still compute the consistent policy. The results are depicted in Figure 4 considering parameters $\omega_{m,1} \in \{0.2, 0.4, 0.6, 1, 1.5, 2, 2.5, 3, 3.5\}$ and $\delta = 0$.

VII. Conclusion

We have proposed a new framework to determine when a partially observable Markov decision process with finite state, input and measurement spaces should be sampled or/and intervened, assuming these operations are costly. We proposed two approaches to find a policy for the time intervals between interventions based on available data in between these interventions, in a event-triggered control fashion. The approach based on relaxed dynamic programming can provide nearly optimal cost when the state dimension is very small, but it is not suitable for larger dimensional problems, as this involves solving large linear programs whose dimension increases significantly with the state dimension. In turn the approach based on consistent control provides a simple and effective solution to the problem.

References

[1] G. Vellidis, M. Tucker, C. Perry, C. Kvien, and C. Bednarz, "A real-time wireless smart sensor array for scheduling irrigation," *Comput. Electron. Agric.*, vol. 61, no. 1, p. 44–50, apr 2008.

[2] J. A. Cabrera, J. R. Pedrasa, A. M. Radanielson, and A. Aswani, "Mechanistic crop growth model predictive control for precision irrigation in rice," in *2021 European Control Conference (ECC)*, 2021, pp. 1243–1248.

[3] J. Shanahan, N. Kitchen, W. Raun, and J. Schepers, "Responsive in-season nitrogen management for cereals," *Computers and Electronics in Agriculture*, vol. 61, no. 1, pp. 51–62, 2008, emerging Technologies For Real-time and Integrated Agriculture Decisions.

[4] Q.-L. Li, J.-Y. Ma, R.-N. Fan, and L. Xia, "An overview for markov decision processes in queues and networks," 2019.

[5] L. Xia and Q.-S. Jia, "Parameterized markov decision process and its application to service rate control," *Automatica*, vol. 54, pp. 29–35, 2015.

[6] Y. Ran, X. Zhou, P. Lin, Y. Wen, and R. Deng, "A survey of predictive maintenance: Systems, purposes and approaches," *ArXiv*, vol. abs/1912.07383, 2019.

[7] R. Schöbi and E. N. Chatzi, "Maintenance planning using continuous-state partially observable markov decision processes and non-linear action models," *Structure and Infrastructure Engineering*, vol. 12, no. 8, pp. 977–994, 2016.

[8] R.-B. Xue, X.-S. Ye, and X.-R. Cao, "Optimization of stock trading with additional information by limit order book," *Automatica*, vol. 127, p. 109507, 2021.

[9] Y.-W. Bai, Z.-L. Xie, and Z.-H. Li, "Design and implementation of an embedded home surveillance system with ultra-low alert power," in *2011 IEEE International Conference on Consumer Electronics (ICCE)*, 2011, pp. 295–296.

[10] A. T. J. R. Cobbenhagen, D. J. Antunes, M. J. G. van de Molengraft, and W. P. M. H. Heemels, "Coverage control for outbreak dynamics," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 984–989.

[11] W. Heemels, K. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *IEEE Conference on Decision and Control (CDC) 2012, Hawaii, USA*, December 2012, pp. 3270–3285.

[12] D.-X. Wang and X.-R. Cao, "Event-based optimization for pomdps and its application in portfolio management," in *18th IFAC World Congress Milano (Italy)*, 2011, pp. 3228–3233.

[13] L. Xia, Q. Jia, and X. Cao, "A tutorial on event-based optimization—a new optimization framework," *Discrete Event Dyn. Syst.*, vol. 24, pp. 103–132, 2014.

[14] J. Messias, M. Spaan, and P. Lima, "Multiagent pomdps with asynchronous execution," in *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AA- MAS 2013)*, 2013, pp. 1273–1274.

[15] A. Molin and S. Hirche, "Structural characterization of optimal event-based controllers for linear stochastic systems," in *Proceedings of the IEEE International Conference on Decision and Control (CDC 2010)*, 2010, pp. 3227–3233.

[16] Y. Xu and J. Hespanha, "Optimal communication logics in networked control systems," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 4, Dec 2004, pp. 3527–3532 Vol.4.

[17] D. J. Antunes and M. H. Balaghi I., "Consistent event-triggered control for discrete-time linear systems with partial state information," *IEEE Control Systems Letters*, vol. 4, no. 1, pp. 181–186, Jan 2020.

[18] D. J. Antunes and B. A. Khashooei, "Consistent dynamic event-triggered policies for linear quadratic control," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1386–1398, 2018.

[19] V. Krishnamurthy and D. V. Djonin, "Structured threshold policies for dynamic sensor scheduling—a partially observed markov decision process approach," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4938–4957, 2007.

[20] V. Krishnamurthy, "Pomdp sensor scheduling with adaptive sampling," in *17th International Conference on Information Fusion (FUSION)*, 2014, pp. 1–7.

[21] M. Rezaeian, "Sensor scheduling for optimal observability using estimation entropy," in *Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07)*, 2007, pp. 307–312.

[22] B. Lincoln and A. Rantzer, "Relaxing dynamic programming," *IEEE Trans. on Automatic Control*, vol. 51, no. 8, pp. 1249–1260, Aug 2006.

[23] A. O. Hero, D. A. Castanon, D. Cochran, and K. Kastella, *Foundations and Applications of Sensor Management*, 2008.

[24] S. I. Resnick, *Adventures in stochastic processes*. Basel, Switzerland, Switzerland: Birkhauser Verlag, 1992.

[25] G. Grimmett and D. Stirzaker, *Probability and random processes (3rd ed.)*. Oxford University Press, 2001.

VIII. PROOFS

A. Proof of Lemma 1

The first statement follows directly from Assumptions 1 and 3 and the structure of the model. Due to these assumptions, the process state has the same probability distribution at s_ℓ when (5a) holds or at $s_{\ell+1}$ when (5b) holds. Moreover, the model can be written in terms of the functions f, g, and h given in Section II-A, which only depend on the process variables in an epoch $\{s_\ell, s_\ell+1, \dots, s_{\ell+1}\}$. This allows one to conclude that indeed the stopping time problems (18) are identical and admit a solution that only depends on the information set $\mathcal{H}_t^\ell$, as shown in (19).

To prove the second statement, we consider an arbitrary policy for the stopping times τ_ℓ taking the form (19) but not necessary optimal. We start by noticing that due to the decomposition of J_s given in (31), we can also decompose J_{av} as

$$J_{\text{av}} = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{\ell=0}^{L(T)-1}\theta_\ell + \sum_{t=s_{L(T)}}^{T-1} g(x_t,\zeta_t,\phi_{L(T)})\right], \qquad (31)$$

with

$$\theta_\ell = \sum_{t=s_\ell}^{s_{\ell+1}-1} g(x_t,\zeta_t,\phi_\ell) + \delta$$

independently and identically distributed random variables due to the first part of the present lemma. When taking the limit, the last term vanishes and due to Wald's identity [24, Ch.1]

$$\mathbb{E}[\sum_{\ell=0}^{L(T)-1}\theta_\ell] = \mathbb{E}[L(T)]\mathbb{E}[\theta_\ell]$$

where

$$\mathbb{E}[\theta_\ell] = \left(\mathbb{E}[\sum_{t=s_\ell}^{s_\ell+\tau_\ell-1} g(x_{s_\ell+t},\zeta_{s_\ell+t},\phi_{s_\ell})] + \delta\right)$$

Moreover, due to the key renewal theorem [24, Ch.3]

$$\lim_{T\to\infty}\mathbb{E}[L(T)] = \frac{1}{\mathbb{E}[\tau_\ell]}$$

From these equations we conclude that for an arbitrary policy for τ_ℓ taking the form (19)

$$J_{\text{av}} = \frac{1}{\mathbb{E}[\tau_\ell]}\left(\mathbb{E}[\sum_{t=s_\ell}^{s_\ell+\tau_\ell-1} g(x_{s_\ell+t},\zeta_{s_\ell+t},\phi_{s_\ell})] + \delta\right)$$

Thus, the optimal policy that minimizes the right-hand side is also an optimal policy for J_{av}

In order to prove the third statement of the present Lemma, note that if the optimal cost (18) of the optimal policy is lower bounded by β, $J_{\text{s},\mu} \geq \beta$ then it must be the case that, for any policy for τ,

$$\mathbb{E}[\sum_{t=s_\ell}^{s_\ell+\tau_\ell-1} g(x_{s_\ell+t},\zeta_{s_\ell+t},\phi_{s_\ell})] + \delta \geq \beta\mathbb{E}[\tau_\ell]$$

or equivalently that

$$\mathbb{E}[\sum_{t=s_\ell}^{s_\ell+\tau_\ell-1} (g(x_{s_\ell+t},\zeta_{s_\ell+t},\phi_{s_\ell}) - \beta)] + \delta \geq 0$$

for any policy. Conversely for some β if we can find a policy for which

$$\mathbb{E}[\sum_{t=s_\ell}^{s_\ell+\tau_\ell-1} (g(x_{s_\ell+t},\zeta_{s_\ell+t},\phi_{s_\ell}) - \beta)] + \delta \leq 0$$

then $J_{\text{av}} \leq \beta$. Since the number of policies for τ is finite, say P, we can write the optimal cost as the minimum over a finite number of continuous function of β (for fixed policy)

$$\mathbb{E}[\sum_{t=s_\ell}^{s_\ell+\tau_\ell-1} (g(x_{s_\ell+t},\zeta_{s_\ell+t},\phi_{s_\ell}) - \beta)] + \delta$$

and thus the cost is a continuous function of β. Moreover, it is clear that it is non-increasing with β and that it is positive when $\beta = 0$ and negative for very large β. The value of β that leads to zero cost is thus equal to $J_{av} = \beta$ and the corresponding optimal policy is an optimal policy for the original problem concluding the proof.

B. Proof of Lemma 2

Using induction, we have $J_{\bar{h}}(q_{\bar{h}}) = V_{\bar{h}}(q_{\bar{h}}) = \delta$ and assuming

$$J_{t+1}(q_{t+1}) \leq V_{t+1}(q_{t+1}) \leq J_{t+1}(q_{t+1}) + \epsilon(\bar{h}-(t+1)),$$

then

$$\begin{aligned}
\mathcal{J}_t^\epsilon(q_t) &= \min\{\delta+\epsilon, (q_t^\intercal\bar{g} - \beta + \epsilon) + \mathbb{E}[V_{t+1}(q_{t+1})|\mathcal{H}_t]\}\\
&\leq \min\{\delta+\epsilon, (q_t^\intercal\bar{g} - \beta + \epsilon) + \mathbb{E}[V_{t+1}(q_{t+1})|\mathcal{H}_t]\}\\
&\leq \min\{\delta+\epsilon(\bar{h}-t), (q_t^\intercal\bar{g} - \beta + \epsilon) +\\
&\qquad \mathbb{E}[J_{t+1}(q_{t+1}) + \epsilon(\bar{h}-(t+1))|\mathcal{H}_t]\}\\
&= \epsilon(\bar{h}-t) + \min\{\delta, (q_t^\intercal\bar{g} - \beta + \mathbb{E}[J_{t+1}(q_{t+1})|\mathcal{H}_t]\}\\
&= J_t(q_t) + \epsilon(\bar{h}-t)
\end{aligned}$$

The relaxed dynamic programming algorithm ensures through (24) that

$$V_t(q_t) = \min_{c\in\mathcal{V}_t} c^\intercal q \leq \mathcal{J}_t^\epsilon(q_t)$$

from which we conclude the desired inequality

$$V_t(q_t) \leq J_t(q_t) + \epsilon(\bar{h}-t).$$

(The inequality $J_t(q_t) \leq V_t(q_t)$ is clear).

C. Proof of Lemma 3

Let us first assume that (5a) holds. From the proof of Lemma 1 we conclude that for any policy for the stopping times τ_ℓ and for an arbitrary $\ell \in \mathbb{N}_0$

$$J_{\text{av}} = \frac{1}{\mathbb{E}[\tau_\ell]}\left(\mathbb{E}[\sum_{t=s_\ell}^{s_\ell+\tau_\ell-1} g(x_{s_\ell+t},\zeta_{s_\ell+t},\phi_{s_\ell})] + \delta\right)$$

which specialized to $\tau_\ell = h$ and $\ell = 0$ ($s_\ell = 0$) leads to

$$J_{\text{av}} = \frac{1}{h}\left(\mathbb{E}[\sum_{t=0}^{h-1} g(x_t,t,\phi_0)] + \delta\right) \qquad (32)$$

Note that

$$\mathbb{E}[\sum_{t=0}^{h-1} g(x_t,t,\phi_0)] = \sum_{i=1}^{b} a_i \mathbb{E}[\sum_{t=0}^{h-1} g(x_t,t,i)]$$

Due to (5a),

$$\mathbb{E}[g(x_0,0,i)] = \bar{g}_0^{\mathsf{T}} \rho_i$$

and since the probability distribution of x_t (when measurements y_t are unknown) is simply $P^t\rho_i$) we obtain

$$\mathbb{E}[g(x_t,t,i)] = \bar{g}_t^{\mathsf{T}} P^t \rho_i$$

Replacing these expressions in (32) we obtain $J_{\text{av}} := J_s + \delta\frac{1}{h}$ with J_s given by (26) as desired.

Let us now assume that (5b) holds. In the first epoch $\{s_0, s_0+1, \ldots, s_1-1\}$ the cost will be slightly different than in the subsequent epochs since the distribution of $x_0 = x_{s_0}$ is different from those of x_{s_ℓ}, which are all the same for $\ell \geq 1$ due the reset imposed at times $s_\ell+1$ imposed by (5b). Since we are interested in an average cost this plays no role and we can simply compute (32) considering the epoch $\{s_1, s_1+1, \ldots, s_2-1\}$ with $s_1 = h$ and $s_2 = 2h$, i.e.,

$$J_{\text{av}} = \frac{1}{h}\left(\mathbb{E}[\sum_{t=h}^{2h-1} g(x_t, t-h, \phi_1)] + \delta\right)$$

Since the distribution of x_1 is known to be ρ_i with $i = \phi_1$, the distribution of x_h is $P^{t-1}\rho_i$. The distribution of x_{h+1} is ρ_i and of x_{h+1+r}, $r \in \{1,2,\ldots,h-1\}$ is $P^r\rho_i$. Thus, in this case

$$\mathbb{E}[g(x_{h+t},t,i)] = \bar{g}_t^{\mathsf{T}} P^{t-1}\rho_i, t \in \{1,\ldots,h-1\}$$

and

$$\mathbb{E}[g(x_h,0,i)] = \bar{g}_0^{\mathsf{T}} P^{h-1}\rho_i$$

Replacing these expressions in (32) we obtain $J_{\text{av}} := J_s + \delta\frac{1}{h}$ with J_s given by (26) as desired.

D. Proof of Theorem 1

Due to Lemma 1 it suffices to consider $\ell = 0$ and for simplicity we consider $b = 1$. The proposed policy in this interval can be rewritten as follows:

$$\tau_0 = \min\{r \in \{1,\ldots,\bar{h}\} | \sum_{t=0}^{r} \mathbb{E}[g(x_t,t,1)|\mathcal{H}_t^0] > -\delta + \omega_{c,1} r\} \tag{33}$$

This ensures that

$$\sum_{t=0}^{\tau_0-1} \mathbb{E}[g(x_t,t,1)|\mathcal{H}_t^0] + \delta - \omega_{c,1} s \leq 0$$

Thus

$$\mathbb{E}[\sum_{t=0}^{\tau_0-1} \mathbb{E}[g(x_t,t,1)|\mathcal{H}_t^0] + \delta - \omega_{c,i} s] \leq 0$$

We will prove that

$$\mathbb{E}[\sum_{t=0}^{\tau_0-1} \mathbb{E}[g(x_t,t,1)|\mathcal{H}_t^0]] = \mathbb{E}[\sum_{t=0}^{\tau_0-1} g(x_t,t,1)] \tag{34}$$

so that the inequality above implies

$$\frac{1}{\mathbb{E}[\tau_0]}\mathbb{E}[\sum_{t=0}^{\tau_0-1} g(x_t,t,1)] + \delta\frac{1}{\mathbb{E}[\tau_0]} \leq \omega_{c,1} < \omega_{m,1}$$

leading to the desired consistency conclusion.

It suffices to establish (34). To this effect we start by noticing that

$$\eta_r := \sum_{t=0}^{r-1} (\mathbb{E}[g(x_t,t,1)|\mathcal{H}_t^0] - g(x_t,t,1))$$

with η_0 is a martingale with respect to the filtration $\mathcal{H}_t^0$ since $\mathbb{E}[\eta_{j+1}|\mathcal{H}_j^0] = \eta_j + \mathbb{E}[\mathbb{E}[g(x_j,j,1)|\mathcal{H}_j^0] - g(x_j,j,1)|\mathcal{H}_j^0] = \eta_j$ Due to the optional sampling theorem [25, Sec. 12.4, Th. 11] [19,] (which can be applied since $\tau_0 \bar{h}$)

$$\mathbb{E}[\eta_{\tau 0}] = 0$$

so that

$$\mathbb{E}[\sum_{t=0}^{\tau_0-1} \mathbb{E}[g(x_t,t,1)|\mathcal{H}_t^0] - (\sum_{t=0}^{\tau_0-1} g(x_t,t,1))] = 0$$

which implies (34).